

AI and the Everything in the Whole Wide World Benchmark

Year: 2021 | Citations: 386 | Authors: Inioluwa Deborah Raji, Emily M. Bender, Amandalynne Paullada, Emily L. Denton, A. Har

Abstract

There is a tendency across different subfields in AI to valorize a small collection of influential benchmarks. These benchmarks operate as stand-ins for a range of anointed common problems that are frequently framed as foundational milestones on the path towards flexible and generalizable AI systems. State-of-the-art performance on these benchmarks is widely understood as indicative of progress towards these long-term goals. In this position paper, we explore the limits of such benchmarks in order to reveal the construct validity issues in their framing as the functionally "general" broad measures of progress they are set up to be.